[苏黎世大学在 Reddit 搞 AI 实验，用虚假身份引伦理争议，网友怒了](https://mp.weixin.qq.com/s?__biz=Mzk0OTY1MDkwOQ==&mid=2247487111&idx=5&sn=b4ea268cf29c821ab203f05a875aa3c7)

原创sleuthResearch Integrity2025-04-29 22:08:54新加坡



Research Integrity

近日，苏黎世大学一项在 Reddit 论坛 r/ChangeMyView（CMV）板块进行的 AI 实验，因伦理问题引发广泛争议。该研究旨在探究大语言模型（LLMs）能否改变人们的观点，却因未获参与者知情同意而遭到猛烈批评。



4 月早些时候，CMV 版主发现来自苏黎世大学的研究人员未经授权，在论坛上使用 AI 生成评论。这些评论以多种虚假身份发布，包括强奸受害者、创伤咨询师、反对 “黑人的命也是命” 运动的黑人等，违反了论坛禁止 AI 生成内容及误导性冒充他人的规则。该研究成果初步报告已由苏黎世大学研究团队分享。

周六清晨，CMV 版主发布长文，详细说明了此次未经授权的实验。研究人员虽匿名向版主提供研究描述，称过去数月用多个账号在 CMV 发帖，评估 LLM 在特定伦理场景中的说服力，且人工审核评论确保无害，但未向用户披露 AI 使用情况，因担心影响研究可行性。

这一实验随即引发网友强烈谴责。科罗拉多大学信息科学家 Casey Fiesler 称，这是其见过最严重的研究伦理违规行为之一，未经同意欺骗性地操纵在线社区用户并非 “低风险”，已造成伤害。康奈尔大学公民与技术实验室研究主任 Sara Gilbert 则指出，该研究损害了 CMV 社区本身，人们可能不再信任交流对象是否为机器人，社区使命或受影响。

对此，苏黎世大学媒体关系办公室暂未回应相关询问。不过，该校文理学院伦理委员会已调查此事，并向首席研究员发出正式警告。此前，CMV 版主已向学校机构审查委员会投诉，并要求阻止研究发表，但学校认为该项目有重要意义，研究产生的风险极小，压制发表与研究成果重要性不相称。

CMV 版主 u/DuhChappers 表示，实验违反了 Reddit 相关规则，研究人员未事先沟通且主动操纵社区成员，越过了不可接受的界限。此前，OpenAI 也曾使用 CMV 内容测试 AI 模型说服力，但未对未同意的人类对象进行实验。

此次事件引发公众对科研伦理及 AI 应用边界的深入思考，也给未来相关研究敲响警钟。

https://retractionwatch.com/2025/04/28/experiment-using-ai-generated-posts-on-reddit-draws-fire-for-ethics-concerns/

**来源：公众号Research Integrity，转载请注明出处，若没注明学术诚信公众号出处，构成侵权。后台联系客服微信：BikElisabeth**

免责声明：

质疑信息来源于Pubpeer，提及人名均为音译

对于文章内容的真实性、完整性、及时性

本公众号不做任何保证或承诺，仅供读者参考

未经授权禁止转载！

转载请勿更改原文内容及格式！

如有转载需求或合作事宜

可添加下方客服微信或推送邮件到researchintegrity@qq.com

